Class5 homework

AUTHOR
(SungWoo Park PID:A69026846)

A. Can you improve this analysis code?

df <- data.frame(a=1:10, b=seq(200,400,length=10),c=11:20,d=NA)

```
df <- data.frame(a=1:10, b=seq(200,400,length=10),c=11:20,d=NA)
df$a <- (df$a - min(df$a)) / (max(df$a) - min(df$a))
df$b <- (df$b - min(df$a)) / (max(df$b) - min(df$b))
df$c <- (df$c - min(df$c)) / (max(df$c) - min(df$c))
df$d <- (df$d - min(df$d)) / (max(df$a) - min(df$d))
```

```
Answer_A <- function(x) {
  ( x - min(x, na.rm=TRUE)) / (max(x, na.rm=TRUE) - min(x, na.rm=TRUE))
}
```

```
Answer_A(df$a)
```

```
 [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
 [8] 0.7777778 0.8888889 1.0000000
```

B. Next improve the below example code for the analysis of protein drug interactions

abstracting the main activities in your own new function. Then answer questions 1 to 6 below. It is recommended that you start a new Project in RStudio in a new directory and then install the bio3d package noted in the R code below

```
library(bio3d)
s1 <- read.pdb("4AKE") # kinase with drug
```

```
  Note: Accessing on-line PDB file
```

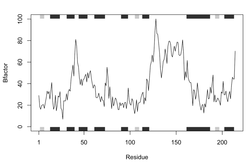
```
s2 <- read.pdb("1AKE") # kinase no drug
```

```
  Note: Accessing on-line PDB file
    PDB has ALT records, taking A only, rm.alt=TRUE
```

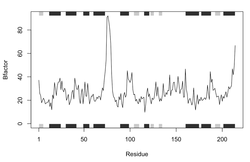
```
s3 <- read.pdb("1E4Y") # kinase with drug
```

```
  Note: Accessing on-line PDB file
```

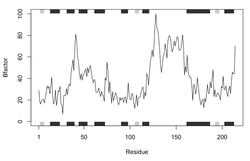
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s1, chain="A", elety="CA")
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```




```
plotb3(s2.b, sse=s2.chainA, typ="l", ylab="Bfactor")
```




```
plotb3(s3.b, sse=s3.chainA, typ="l", ylab="Bfactor")
```



#Q1 What type of object is returned from the read.pdb() function?

This function code returns a list of class 'pdb' from protein data bank, representing the protein structure. Value (atom, helix, sheet, seqres, xyz, calpha, remark, call).

#Q2. What does the trim.pdb() function do?

This function is for trimming the protein structure and focus on specific chain of protein or type of atoms.

#Q3. What input parameter would turn off the marginal black and grey rectangles in the plots and what #do they represent in this case?

Black rectangles and grey rectangles represent the secondary structure elements (SSEs) in the protein structure. If we change the sse argument in plotb3() to FALSE, as in the following example: plotb3(s1.b, sse = FALSE, typ = "l", tlab = "Bfactor", then the rectangles will disappear.

#Q4. What would be a better plot to compare across the different proteins?

I think scatter plot will be better to compare B-factors across different proteins.

#Q5. Which proteins are more similar to each other in their B-factor trends? How could you quantify this? s1 and s3 is more closer.

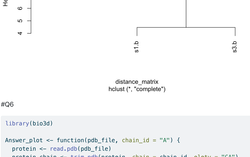
```
s1.b <- c(10, 20, 30, 40, 50)
s2.b <- c(15, 25, 35, 45, 55)
s3.b <- c(12, 22, 32, 42, 52)

bfactors_matrix <- rbind(s1.b, s2.b, s3.b)

distance_matrix <- dist(bfactors_matrix)

hc <- hclust(distance_matrix)

plot(hc)
```



#Q6

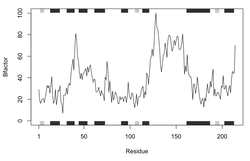
```
library(bio3d)

Answer_plot <- function(pdb_file, chain_id = "A") {
  protein <- read.pdb(pdb_file)
  protein_chain <- trim.pdb(protein, chain = chain_id, elety = "CA")
  b_factors <- protein_chain$atom$b
  plotb3(b_factors, sse = protein_chain, typ = "l", ylab = "Bfactor")
}

Answer_plot("4AKE")
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/qm/pns4b6q539g78xhnryt0vkmm0000gn/T//RtmpQRUw7Z/4AKE.pdb exists.
Skipping download
```

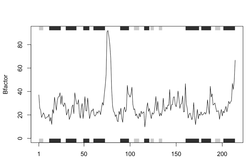



```
Answer_plot("1AKE")
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/qm/pns4b6q539g78xhnryt0vkmm0000gn/T//RtmpQRUw7Z/1AKE.pdb exists.
Skipping download

   PDB has ALT records, taking A only, rm.alt=TRUE
```




```
Answer_plot("1E4Y")
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/qm/pns4b6q539g78xhnryt0vkmm0000gn/T//RtmpQRUw7Z/1E4Y.pdb exists.
Skipping download
```